Sound properties associated with equiluminant colours

Article (Accepted Version)

Hamilton-Fletcher, Giles Douglas Iain, Witzel, Christoph, Reby, David and Ward, Jamie (2017) Sound properties associated with equiluminant colours. Multisensory Research, 30 (3-5). pp. 337-362. ISSN 2213-4794

Sound Properties Associated with Equiluminant Colours

Giles Hamilton-Fletcher[1,2,*], Christoph Witzel[3], David Reby[1] and Jamie Ward[1,2]

[1]School of Psychology, University of Sussex, Brighton, UK

[2]Sackler Centre for Consciousness Science, University of Sussex, Brighton, UK

[3]Allgemeine Psychologie, Justus-Liebig-Universität Gießen, Gießen, Germany

Received 30 September 2016; accepted 27 March 2017

* To whom correspondence should be addressed. E-mail: gh71@sussex.ac.uk

Abstract

There is a widespread tendency to associate certain properties of sound with those of colour (e.g., higher pitches with lighter colours). Yet it is an open question how sound influences chroma or hue when properly controlling for lightness. To examine this, we asked participants to adjust physically equiluminant colours until they 'went best' with certain sounds. For pure tones, complex sine waves and vocal timbres, increases in frequency were associated with increases in chroma. Increasing the loudness of pure tones also increased chroma. Hue associations varied depending on the type of stimuli. In stimuli that involved only limited bands of frequencies (pure tones, vocal timbres), frequency correlated with hue, such that low frequencies gave blue hues and progressed to yellow hues at 800 Hz. Increasing the loudness of a pure tone was also associated with a shift from blue to yellow. However, for complex sounds that share the same bandwidth of frequencies (100–3200 Hz) but that vary in terms of which frequencies have the most power, all stimuli were associated with yellow hues. This suggests that the presence of high frequencies (above 800 Hz) consistently yields yellow hues. Overall we conclude that while pitch–chroma associations appear to flexibly re-apply themselves across a variety of contexts, frequencies above 800 Hz appear to produce yellow hues irrespective of context. These findings reveal new sound–colour correspondences previously obscured through not controlling for lightness. Findings are discussed in relation to understanding the underlying rules of cross-modal correspondences, synaesthesia, and optimising the sensory substitution of visual information through sound.

Keywords

Correspondences, vision, colour, hearing, sound

1. Introduction

In the opening moments of Disney's 1940 animated production of *Fantasia*, an abstract visual interpretation of the sounds of the Philadelphia Orchestra playing Bach's staccato fugue is created for the viewer. The abstract images vary in colour, shape and location and are the results of visual artists' mental imagery. The full-bodied steady brass section gives rise to all-encompassing swathes of deeply saturated reds and oranges, while the high-pitched string instruments give rise to smaller angular columns of bright light with rapid changes in notes reflecting fast motion across the screen, tending to rise with ascending pitch or fall with descending. Are these associations arbitrary and idiosyncratic, or do they reveal more general patterns of associations across modalities?

In experimental studies, variations of lightness, location, shape, and chroma have been associated to changes in pitch, loudness, tempo and timbral qualities (Marks, 1974; Walker, 2012; Walker *et al*., 2010; Ward *et al*., 2006). These intuitive pairings across the senses are referred to as cross-modal correspondences in the general population and congruency effects from this appear to influence their aesthetic appeal, integration and perceptual processing (Spence, 2011; Ward *et al*., 2008). Correspondences are often contrasted with developmental synaesthesia, where stimulation in one modality (e.g., auditory) can elicit automatic, consistent and conscious percepts in a second modality (e.g., vision) in a small portion of the population (Novich *et al*., 2011; Simner, 2012a; Simner *et al*., 2006). It should be noted that correspondences and synaesthesia may share some of the same tendencies (e.g., pitch lightness, Ward *et al*., 2006; for a review see Simner, 2013) and as a result it is likely that correspondences (especially those present in infancy) may influence any development of synaesthesia in the related modalities (Ludwig and Simner, 2013; Simner and Ludwig, 2012; Walker *et al*., 2010). In cases of visual deprivation, cases of acquired synaesthesia seem to be most commonly manifested as auditory to visual synaesthesias, suggesting strong predispositions for these regions to connect (Afra *et al*., 2009). Related to this, in sensory substitution devices where visual information is systematically encoded in sound, visually-deprived long term users of such devices have reported a 'visual' phenomenology resulting from sound, as a kind of artificially acquired audio-visual synaesthesia (Ward and Meijer, 2010; Ward and Wright, 2014).

Examining these multi-modal interactions allow us to identify rule sets that the brain uses to pair seemingly separate stimuli together. By examining the nature of these mappings, it becomes possible to determine which neural processes are likely to facilitate such bindings; such as whether these are driven by lower-level features and neural-encoding, for instance, increasing loudness and increasing lightness are associated (Marks, 1987) and both properties are associated by increased neural activation in primary but not secondary cortices (Goodyear and Menon, 1998; Mulert *et al*., 2005); through learned association, such as high-pitch and small objects (Evans and Treisman, 2010) where smaller animals tend have higher pitched voices (Fitch, 1997); higher-level cognitions (e.g., *high* pitch and *high* elevation, sharing the linguistic term ‘high’) including mediating pathways such as emotional valence, where emotionally-positive sounds and colours become associated (Palmer *et al*., 2013; Sebba, 1991). Spence (2011) describes these types as 'structural,' when they manifest as a result of typical brain development, 'statistical' when they are learned from continual exposure to matching stimuli and 'semantic' when they result from a third mediating process, such as language or emotion. Furthermore it is suggested that while the first two affect lower-level perceptual processing, all three can influence higher-level decisions. However a given correspondence might fit into all three categories, such as with pitch height, which have shown a matching preference in pre-verbal infants (Braaten, 1993), an effect on perceptual processing tasks (Evans and Treisman, 2010), feature as statistical correlations in our environment, are exaggerated by the ear's structure (Parise *et al*., 2014), and finally have a stronger effect when the association is reinforced by language (Dolscheid et a.l, 2013).

1.1. Pure Tone Correspondences

The structurally simplest form of sound is a pure tone sine wave, which has a frequency (determining the perceived ‘pitch’) and amplitude (determining the perceived ‘loudness’). These two dimensions interact, so that changing the frequency also subtly changes the perceived loudness. This perceptual phenomenon has been mapped out in loudness-equalisation curves (Fletcher and Munson, 1933; ISO, 2003), where 'phons' is used to describe the perceptual loudness of a given pure tone. Pure tones are thus ideal for examining the effect that one specific frequency has on colour-matching in isolation, and varying their frequency and loudness has revealed a wide variety of colour correspondences.

Increasing the frequency of pure tones is preferentially matched to colours with increased lightness (Ward *et al*., 2006). This correspondence can affect perceptual processing speeds (Marks, 1987; Martino and Marks, 1999; Melara, 1989), is present in children (Mondloch and Maurer, 2004) and is also shared by non-human primates (Ludwig *et al*., 2011), suggesting that this correspondence occurs at the structural level (Spence, 2011). Increasing the loudness of pure tones also relates to increased lightness in both children and adults (Bond and Stevens, 1969; Marks, 1987; Root and Ross, 1965; Stevens and Marks, 1965). However, it is not quite as uniform as the pitch–lightness mappings, as for a smaller subsample of individuals, increasing loudness can be more intuitively put with *decreasing* lightness (Marks, 1974).

The relationship between pure tones and other colour dimensions, such as chroma and hue, is less clear. Ward *et al*. (2006) observed a quadratic relationship between chroma and increasing frequency, whereby frequencies near middle C (262 Hz) would derive the most saturated colours. However, it is likely that this also reflects a pitch–lightness association as it is impossible for extremely dark or bright colours to also be highly saturated, whereas colour with a moderate lightness can potentially be highly saturated. One would therefore expect to see this distribution if participants chose a random selection of chromas and hues, yet still maintained their pitch–lightness tendencies. Giannakis (2001) reports a loudness–chroma mapping for pure tones between 110–3520 Hz, where louder sounds correspond to more saturated colours. This provides an alternative explanation for Ward *et al*.'s inverted U-distribution between frequency and chroma, as extremely high or low pitched pure tones can sound quieter to the listener when not loudness-equalised (ISO, 2003).

Pitch–hue correspondences have been reported in children, placing the highest pitched tones with yellow/green, middle pitches with red/orange and the lowest with blue/purple (Simpson *et al*., 1956). Giannakis (2001) also found that non-synaesthetic participants would combine red/yellow with higher frequencies, green/cyans with middle frequencies, and blue/magenta with the lowest frequencies. In regards to sound–colour synaesthetes, Orlandatou (2012) describes an experiment finding that higher pitched sounds (both pure tone and complex) result in more saturated colours than low-pitched, and that these have a tendency towards yellowish hues. An important note in the direct comparison of these studies is the different frequency ranges used in each experiment, with Simpson *et al*.'s experiment ranging between 125 and 12000 Hz stimuli, Giannakis' experiment is between 110 and 3520 Hz and Orlandatou's is 50–3000 Hz. These studies support an association between high pitches and yellow hues; however, it is unknown whether specific frequencies, or simply the highest frequency in a given context, are associated with yellow. Moreover, the lack of controlling for lightness in these studies means that the reported hue associations might be primarily driven by pitch–lightness associations, with participants secondarily picking prototypical colour exemplars, since yellow is the brightest focal colour (Spence, 2011). Prior studies have also had difficulties in analysing hue in a perceptually meaningful way, either due to using non-human models of colour space, or being unable to meaningfully analyse circular representations of hue (Thornley Head, 2006; Ward *et al*., 2006).

1.2. Complex Sound Correspondences

Increasing auditory complexity beyond simple pure tone stimuli to richer timbral sounds has some important influences on colour selection. Timbre refers to any distinctive qualities in a sound separate from its pitch or loudness. Complex periodic sounds are composed of multiple pure tones and typically consist of a fundamental frequency which is the lowest frequency in a sound (perceived as the sound’s ‘pitch’) and multiples of this frequency (its 'harmonics'). Many factors can be considered when classifying these sounds, such as the range of frequencies used (the sound’s 'bandwidth') and the power distribution for frequencies within the sound (the sound’s 'centre of gravity'). While these factors can be applied to sounds that do not vary over time (called static timbre), there are additional properties for sounds that do vary over time (called dynamic timbre).

Ward *et al*. (2006) found that instrumental sounds (e.g., piano sounds) were given more saturated colours in comparison to pure tone stimuli when played at the same note for both controls and synaesthetes. The increased saturation of instruments in comparison to pure tone stimuli could be explained either through the overall increase in bandwidth or through the density of harmonics that creates a richer sound. Also in this context, the lightness of colours chosen for piano or string instruments were not statistically lighter on average than their pure tone counterparts which points to the fundamental frequency having a key role in influencing the lightness of colours. However, in a different context, when ten timbres were compared all playing the same note, significant differences were found between instrument type and the associated lightness and chroma. This suggests that the distribution and pattern of frequencies beyond the fundamental may influence associated colours' lightness and chroma. Timbre–lightness ranged from the didgeridoo (darkest) to the harp (brightest), while timbre–chroma ranged from the least saturated (didgeridoo/harp) to the most saturated (super tenor/guitar). As stated previously, because minimum and maximum lightness have constraining effects on the potential choice of chroma, it is perhaps not surprising to see the most de-saturated colours occur with the harp and didgeridoo stimuli. However, the lack of spectral analysis on the instrument sounds themselves (which is further complicated by using dynamic timbre that varies over time) did not allow the explicit identification of the aspects of the sound that explains the ordering of these instruments.

Some recent studies have improved both classifications of complex sounds and the analysis of colour dimensions in perceptually accurate colour spaces like CIELUV (Moos *et al*., 2014). These conditions have helped to find specific associations between auditory dimensions that define vowel stimuli and correlate them to lightness as well as saturation towards particular hues. The use of perceptually accurate colour spaces also allows for the capacity to better control for common confounds, such as variations of lightness (Spence, 2011). Finally, in addition to analysing chroma, through circular analysis, it is also possible to analyse hue directly (Batschelet, 1981; Berens, 2009). Overall, a consideration of all of these factors allows a finer degree of control in finding associations between specific dimensions of sound and colour.

1.3. Hypotheses

In order to better understand the relationship between sound and chroma/hue, participants were asked to adjust an equiluminant colour to 'best match' the sounds they were presented with. For pure tones, we expected to see increases in pitch associated with increases in chromaticity, specifically towards yellow hues (Orlandatou, 2012; Simpson *et al*., 1956), and across a variety of contexts (100–3200 Hz and 440–880 Hz). We also sought to replicate previous loudness–chroma findings in a pure tone stimulus (Giannakis, 2001). By examining different bands of frequencies present in a vocal sound, we expected to see higher frequencies yield yellower hues (Moos *et al*., 2014). To examine the influence of a sound's 'centre of gravity' we presented two types of sounds (complex sine waves and vocals) that had the same range of frequencies, but varied in which frequencies had the highest power. We expected that when the higher frequencies have the highest power, this would also increase chroma towards yellow hues.

2. Method

2.1. Participants

Forty-four students of the University of Sussex (33 women aged 19.8 ± 3.1 years old) were recruited. Observers were either paid in course credits or £5 for participation. Prior to the experiment, participants filled out a pre-screening form, in which they confirmed that they had normal or corrected-to-normal visual acuity, had neither hearing nor colour vision deficiencies, and did not experience sound–colour synaesthesia or any other type of synaesthesia.

2.2. Materials

2.2.1. Apparatus

Colours were displayed on a Dell D1626HT 20-inch CRT monitor, driven by an ATI radeon HD 2400 graphics card with a 32-bit colour resolution, a resolution of 1024 by 768 pixels, and a refresh rate of 100 Hz. Colorimetric specifications were measured with a ColourCAL MKII colorimeter (Cambridge Research Systems Ltd.). The CIE1931 chromaticity coordinates and luminance of the monitor primaries were R = [0.627, 0.343, 11.601], G = [0.281, 0.615, 30.346], and B = [0.151, 0.069, 4.21]. Gamma corrections without bit-loss were applied based on the measured gamma curves of the monitor primaries. Observers looked at the display through a viewing tunnel and from a distance of 1 metre. The experimental measurements were done in a dark room in order to control for the observers' adaptation.

Sounds were outputted using SoundMAX HD audio ESP and heard through HD 497 Sennheiser headphones. Experiments were programmed using Matlab (The

Mathworks, Inc.) with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997).

2.2.2. Colour Stimuli

Colours were presented as discs at the centre of a grey monitor background. Disks had 1.75 inches diameter, which corresponds to 2.55 degrees visual angle at the 1 metre viewing distance. Colours were sampled from an isoluminant plane in CIELUV-space. CIELUV consists of an achromatic lightness dimension L*, and two chromatic dimensions, namely a green-red axis u* and a blue-yellow axis v*. The corresponding polar coordinates of CIELUV represent chroma as radius and hue as azimuth (often abbreviated as Lch, for lightness, chroma and hue). CIELUV colour space accounts for the fact that perceived colours are relative to the observer's adaptation by specifying colours relative to an adapting white-point. The adapting white-point was set as xyY = [0.3101 0.3162 50]. The chromaticities of the grey background corresponded to those of the assumed white-point. The lightness of the background was L* = 60. The definition of the white-point with a higher luminance than the adapting background is done to avoid L* values above 100, when stimuli are lighter than the background, as here (cf. Witzel and Franklin, 2014). The lightness of the stimulus colours was fixed at L* = 65, i.e., slightly lighter than the background. A constant lightness L* implies a constant luminance. Controlled measurements revealed that the rendered luminance varied less than 0.5%. Adjustments of hue changed the CIELUV azimuth, adjustments of chroma changed the CIELUV radius of the disk's colour. Adjustments of chroma were constrained to a maximum of 60 because this is the highest chroma available for each hue within the monitor gamut (cf. Witzel and Franklin, 2014; Forder *et al*., 2014).

2.2.3. Auditory Stimuli

Pure tone frequency — set 1 (100–3200 Hz range). As perceived pitch and loudness are related phenomena, loudness-equalised sine waves of varying frequency were produced. The amplitude of individual sine waves was scaled with reference to 40 phons (subjective measure of equal loudness) on equal-loudness-level contours (ISO, 2003). The frequency and amplitude of the stimuli are as follows; 100 Hz (0.92 amp), 200 Hz (0.68 amp), 400 Hz (0.52 amp), 800 Hz (0.4 amp), 1600 Hz (0.4 amp) and 3200 Hz (0.3 amp). This allows a variation in perceived pitch without obvious changes in subjective loudness to co-occur.

Pure tone frequency — set 2 (440–880 Hz range). In order to gauge whether context is important, another set of frequencies was produced that spanned a shorter frequency range. These consisted of sine wave frequencies taken from a musical octave, consisting of 440, 493, 523, 587, 659, 698, 784 and 880 Hz pure tones. This range was chosen as it was close to the middle two frequencies in the previous condition and so it would not consist of especially ‘high’ or ‘low’ frequencies relative to the previous stimulus. Since there are only minimal changes in loudness across these frequencies no loudness equalisation was applied.

Pure tone loudness. In order to create stimuli that varied in loudness but not pitch, a 40 phons 400 Hz pure tone was created. Three additional stimuli were derived from this with 0.5, 0.25 and 0.1 amplitude proportions. This was done so that subjective loudness would be at normal levels with respect to the frequency condition, half-amplitude, quarter-amplitude and finally a 10th of the amplitude as the quietest stimuli.

Timbral frequency bands. In order to understand the impact that various bands of frequencies have on colour selections for richer timbral sounds, a synthesised vocal sound was created and then band-passed through specific frequency bands. An artificial vocal sound was created using Praat voice synthesis and analysis software (Boersma and Weenink, 2012). A 27-year-old male vocal sound was recorded (fundamental frequency of 83 Hz) and this was used as a reference by Praat for creating a synthesised vocal sound without formants (the spectral peaks of intensity that create vowel sounds). This allowed us to have a distinctive fundamental frequency with an equal distribution of power across all frequency ranges. This base sound was then band-passed through either a 100–200 Hz, 200–400 Hz, 400–800 Hz, 800–1600 Hz or 1600–3200 Hz gate. All sounds had the same implied fundamental frequency of 83 Hz since all frequencies present are multiples of this. As a result, these stimuli allowed us to test individual frequency ranges for richer timbral sounds.

Complex sine wave 'centre of gravity'. In order to understand the influence of increasing the power of either higher or lower frequencies when the range of frequencies remains consistent, a complex sound was produced consisting of 100, 200, 400, 800, 1600 and 3200 Hz sine waves together. This sound either had the lowest or highest three frequencies reduced in dB by 33 or 66%. Alongside the original sound this produced five sounds that vary in their power in low to high frequencies while retaining a 100 to 3200 Hz bandwidth.

Timbral 'centre of gravity'. An artificial vocal sound was created in Praat, using the same vocal reference and procedure as the 'timbral frequency bands' stimuli; however, the base stimuli were then band-passed between 100 and 3200 Hz. From this, frequencies either above or below 600 Hz were reduced in dB by 33 or 66%. Including the original sound, this produced five sounds ranging between 100 and 3200 Hz but varying in the power distribution of low (under 600 Hz) or high (over 600 Hz) frequencies.

2.3. Procedure

Prior to each experiment, the CRT monitor was active for 30 minutes to stabilise colour output. Participants sat at the testing computer and told they would be presented with a series of sounds, and then asked to adjust the colour of a disc to a colour that they felt 'best matched' the current sound. They were walked through the controls for altering the colour of the central disc though changing hue, chroma, their navigation speed, and finally for confirming their selection. Participants were initially given two randomly selected stimuli to practice colour adjustment on, and when they were ready to begin were left in a darkened room to control adaption to the monitor background. During the main task, the six stimulus sets were given a random order, and then completed in turn by the participant. For each stimulus set, participants would first listen to all stimuli within that set in order to control for range effects. Then, they re-listened to these sounds and adjusted the presented colour. The order of stimuli was randomised in both the preliminary listening and the adjustment phase. Each sound stimulus was presented to participants for 1 second; participants could re-listen to the current sound stimulus at any time, though, and were given unlimited time to make their colour choice before confirming their selection. When adjustments for a stimulus set were completed, the next stimulus set was presented, and this process would repeat until adjustments were completed for all stimulus sets. The task took approximately 25 minutes to complete on average.

3. Results

For each stimulus set, the average time and colour-space travelled was taken for each participant. On average across all stimulus sets the average time taken for a colour adjustment was 8.33 seconds (SD 4.6), and the average distance in colour-space travelled in CIELUV space was 46 units (SD 14).

3.1. Chroma

Figure 1a illustrates the relationship between the pitch of pure tones and adjusted chroma for the first, logarithmically scaled set spanning 100–3200 Hz and for the second set spanning 440–880 Hz (in red). To test the relationship between adjusted chroma and pitch, we calculated correlations between frequencies of the tones and the corresponding adjustments of chroma averaged across individuals. In both sets of pure tones, frequency was correlated with adjusted chroma [$r(4) = 0.89$, $p = 0.02$, and $r(6) = 0.98$, $p < 0.001$]. This was still true when combining the measurements for both stimulus sets [$r(12) = 0.78$, $p = 0.001$].

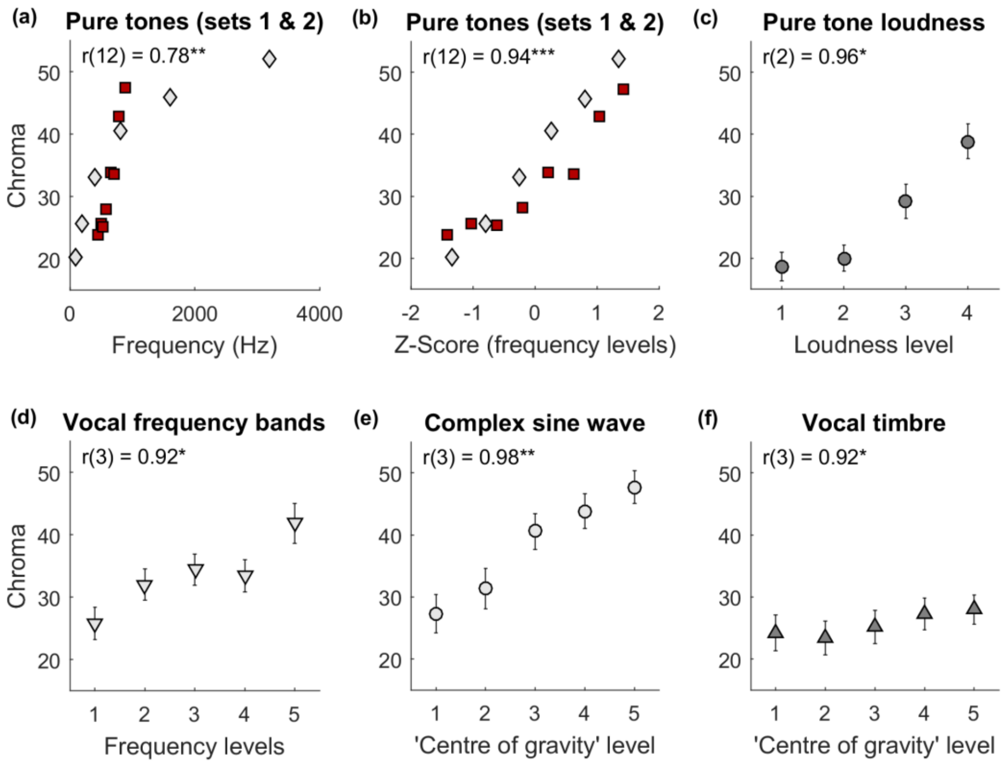

Figure 1. Chroma adjustments for pure tones, complex sine waves and vocal timbres. Panel a illustrates the relationship between the frequency of pure tones for set 1 (grey diamonds) and set 2 (red squares) and chroma average adjustments. Panel b illustrates the same relationship, but with z-scored frequency to account for range effects within each stimulus set. Panel c illustrates results for pure tones that vary in loudness. Panel d illustrates the results for vocal frequency bands that vary in their average frequency. Panels e and f illustrate complex sine wave and vocal timbre sounds that vary in their 'centre of gravity.' Key: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

However, looking at Fig. 1a we realized that the relationship between pitch and chroma is not based on the absolute frequencies; but is instead relative to the range of frequencies in the stimulus set. For example, the first four sounds in the second set (440, 493, 523, and 587 Hz) yield lower chromaticies than the third sound in the first set (400 Hz), a fact that can only be explained if these chroma–pitch mappings are relative to their auditory context. Another point is that the stimuli in the first set follow a logarithmic curve rather than a line, while those in the second set were very close to a line. To account for the relative scaling of frequency, we calculated *frequency levels* according to the rank of stimulus frequency within the set, and we normalized these ranks by z-scores to make them comparable across stimulus sets with different numbers of stimuli. For pure tones, this results in an almost perfect linear relationship between frequency levels and average chroma adjustments when combining the two stimulus sets [$r(12) = 0.94$, $p < 0.001$ — Fig.

1b]. This correlation is also significant after Bonferroni correction for twofold testing of the correlation (α = 0.05/2).

Moreover, to make sure that correlations across only a few stimuli are not spurious, we calculated correlation coefficients for each individual observer, transformed them to Fisher's *z*-transforms (Fisher, 1921), and tested with a *t*-test across observers whether correlation coefficients were larger than zero. Reported average correlation coefficients (*Mr*) were also calculated with Fisher-transforms, and then converted back to *r*-values. Observers who did not adjust chroma at all, were discarded from analyses, which is reflected in the reported degrees of freedom. The *t*-tests confirmed a relationship between the frequency levels of pure tones and chroma [Mr = 0.58, t(43) = 8.8, $p < 0.001$, $d = 1.3$].

Similar results were obtained for the third set of pure tones, which varied in loudness (Fig. 1.c). We calculated the correlation between the average chroma adjustment and the different levels of loudness (1 for low loudness to 4 for high loudness). Chroma adjustments increased with higher loudness levels, only just reaching significance [$r(2) = 0.96$, $p = 0.04$]. However, since we had only four stimuli the degrees of freedom of this correlation is very low. The *t*-test across individuals confirmed a positive correlation between loudness levels and chroma [Mr = 0.86, $t(43) = 3.0$, $p = 0.005$, $d = 0.45$].

As with the pure tones, we found further evidence for the relationship between chroma and frequency with complex sine wave sounds that varied in their 'centre of gravity'. For the complex sine wave sounds (Fig. 1e), we calculated the correlation between the average chroma adjustment and the different 'centre of gravity' levels corresponding to the order of the stimuli (1 for lowest to 5 for highest). Again, we found a high correlation between frequency levels and chroma for average adjustments [$r(3) = 0.98$, $p = 0.002$] and in the *t*-test across individual observers [Mr = 0.68, $t(39) = 5.8$, $p < 0.001$, $d = 0.92$].

Vocal sounds also yielded positive correlations between frequency levels and chroma as observed for the other three stimulus sets. A correlation with chroma was also found for both sets of vocal sounds (Figs 1d, f). For the vocal frequency band stimuli we calculated the correlation between chroma adjustments and their relative frequency level. Positive correlations were found for the analyses of the aggregated data, just reaching significance [$r(3) = 0.92$, $p = 0.026$], but strongly significant for the individual data [*Mr* = 0.42, $t(41) = 3.3$, $p = 0.002$, $d = 0.51$]. For the second set of vocal sounds, we also coded their ‘centres of gravity’ through their relative frequency levels. These frequency levels also just correlated with the aggregated [$r(3) = 0.92$, $p = 0.03$] and individual chroma adjustments [Mr = 0.23, $t(42) = 2.1$, $p = 0.04$, $d = 0.32$]. However, average choma adjustments (Fig. 1f) vary much less across ‘centre of gravity’ levels for vocal sounds than for the other four sound sets varying in frequency levels (Figs 1b, d, e). This suggests that, while being systematic, the effect of frequency on chroma is much weaker for this than for the other sets of sounds.

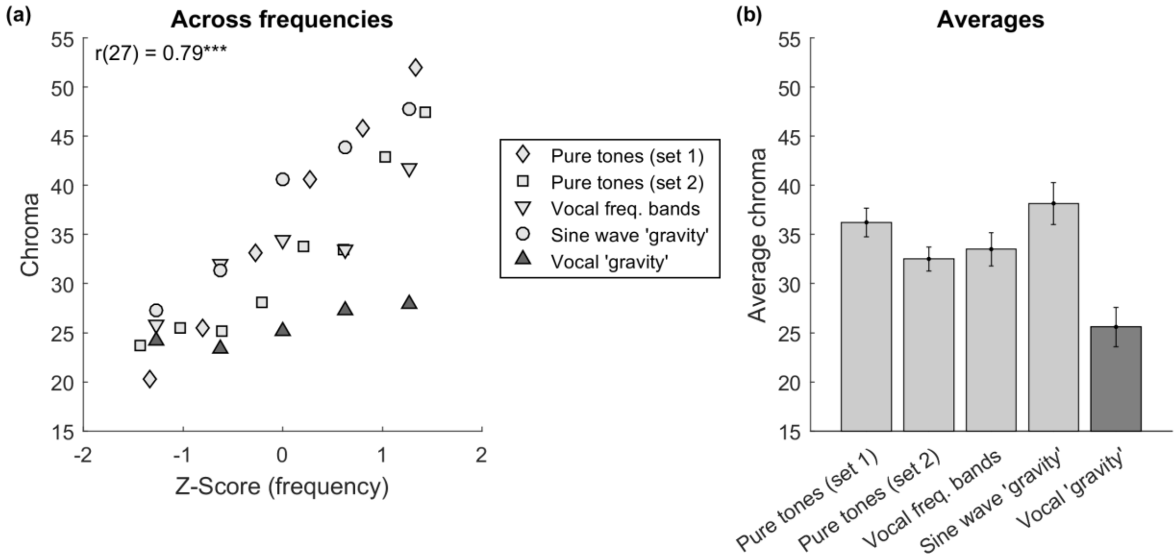

Figure 2. Chroma adjustments for all stimulus sets in which the average frequency varied. Panel a illustrates the relationship between chroma and frequency (z-scored) across all stimulus sets in which the average frequency varied. Panel b illustrates the average chroma adjustments across all stimulus sets in which the average frequency varied. Key: *** $p < 0.001$.

Figure 2 allows for a comparison across all stimulus sets. Panel a illustrates the relationship between chroma and frequency when combining the data from all stimulus sets with z-scored frequencies. The correlation between frequency levels and chroma explains more than 62% of the variance [$r(27) = 0.79$; $p < 0.001$]. A large proportion of unexplained variance is due to the vocal 'centre of gravity' stimuli (highlighted through a darker colour in Fig. 2a). Without that stimulus set, the correlation explains 83% of the variance in the other four stimulus sets [$r(22) = 0.91$, $p < 0.001$]. Note that this correlation is also significant after Bonferroni correction for twofold testing of the correlation ($\alpha = 0.05/2$). t-Tests across individuals also confirmed this relationship [Mr = 0.42, $t(43) = 8.7$, $p < 0.001$, $d = 1.3$].

Finally, Fig. 2b illustrates the average adjustments per stimulus set. For each stimulus set and each observer, we calculated the chroma adjustment averaged across the stimuli of the stimulus set. With the resulting data, we conducted a one-way repeated measures analysis of variance across observers with the five stimulus sets as the factor. Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated, $\chi^2(9) = 20.16$, $p = 0.017$, with $\varepsilon > 0.75$, and therefore, a Huynh–Feldt correction was used. There were significant differences in chroma adjustments across the five stimulus sets [$F(3.57,153.57) = 9.38$, $p < 0.001$, $\eta_p^2 = 0.179$]. Bonferroni-corrected post-hoc *t*-tests revealed that the vocal 'gravity' sounds were significantly less chromatic than the first set of pure tones ($p < 0.001$), vocal frequency bands ($p = 0.014$), sine wave 'gravity' sounds ($p < 0.001$), and not quite significant for the second set of pure tones ($p = 0.058$).

Taken together, average adjustments of chroma differed across stimulus sets. At the same time, all stimulus sets consistently showed a positive relationship between frequency level and chroma, indicating that the higher the frequency of a sound within the range of a stimulus set (*relative pitch*), the more chroma observers associate with the sound.

3.2. Hue

The Circular Statistics toolbox for Matlab Version 2012 (Berens, 2009) was used to analyse hue adjustments. First, we explored central tendencies towards a certain hue in the adjustments of a given stimulus that are useful to better understand the main results reported later on. For this purpose, we used Rayleigh's test for uniformity, which is particularly fit for testing unimodal distributions i.e., tendencies towards one particular hue (Batschelet, 1981; Berens, 2009). We applied this test to the hue adjustments of each stimulus in each stimulus set. Applying a Bonferroni correction for all stimuli in all sets vastly reduces the α level at which they are considered ($0.05/33 = 0.0015$), potentially obscuring important context and interesting tendencies within and across datasets. For this reason, the results of the Rayleigh tests are based on uncorrected significance levels ($\alpha = 0.05$) and should be considered as exploratory.

For the three stimulus sets of pure tones, Rayleigh's tests only had sounds at 1600 Hz and 3200 Hz in the first set and 440 Hz and 880 Hz in the second set, yielding significant tendencies towards one particular hue (mean vector length $V = 0.26$–0.41, *p*s < 0.05). The 440 Hz stimulus was associated with blue hues, while 880

Hz, 1600 Hz and 3200 Hz were associated with yellow-orange hues (see online Supplementary Fig. S1). For the different frequency bands of vocal sounds, two of the five were associated with a specific hue (V = 0.36–0.42, ps < 0.05), an average frequency of 150 Hz was associated with blue hues, while higher average frequencies (2400 Hz) were associated with yellow hues. For sounds with the same bandwidth (100–3200 Hz) that varied in their 'centre of gravity,' three of the five complex sine wave stimuli were significantly associated with one hue (V = 0.30–0.42, ps < 0.05), while all vocal timbre stimuli were associated with one hue (V = 0.27–0.49, ps < 0.05). Both of these sets of stimuli produced yellow hues (Fig. 3c and Supplementary Fig. S2).

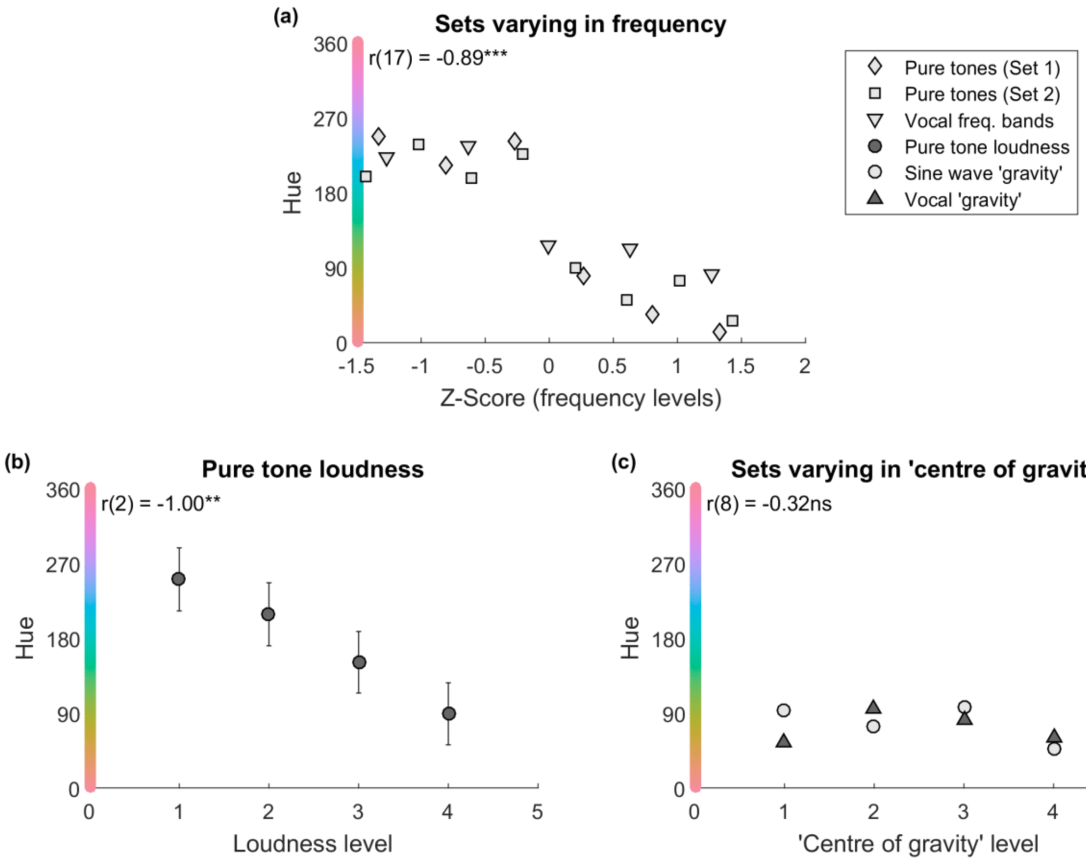

Figure 3. Change of hue with frequency, loudness, and 'centre of gravity'. Statistics in the upper left corners report the correlation across all stimuli varying in frequency (pure tones set 1, set 2, vocal frequency bands) based on Fisher's z-transform across individuals (panel a), different levels of pure tone loudness (panel b), and different levels of 'centre of gravity' (panel c). Key: ** p < 0.01, *** p < 0.001. Note that hue adjustments change linearly as a function of frequencies for pure tones and vocal frequency bands (panel a), and as a function of loudness (panel b); but stay almost constant for sounds that vary in their centre of gravity (panel c). A more detailed breakdown of hue for individual stimuli and sets can be found in Supplementary Figs S1 and S2.

We then examined whether different stimuli in each set yielded different central tendencies in hue adjustments. Because mean vector length was below 0.45

for many of the stimuli, we compared stimuli through a circular median rather than a Watson–Williams test (Berens, 2009). The medians in the first set of pure tones (spanning 100 to 3200 Hz) did not differ significantly across the six tones (*P*-statistic = 8.9, $p = 0.12$). Differences were close to significance in the smaller range of frequencies (440–880 Hz) and those that varied in loudness (*P*-statistic = 13.3, $p = 0.07$ and *P*-statistic = 7.4, $p = 0.06$). There was also no significant difference between the medians of the 'centre of gravity' sounds, whether they were composed of complex sine waves (*P*-statistic = 4.6 $p = 0.33$) or a vocal timbre sounds (*P*-statistic = 3.6, $p = 0.46$). The only stimulus set that yielded clear and highly significant differences across stimuli was the vocal frequency bands (*P*-statistic = 21.2, $p < 0.001$). These results only support the conclusion that hue adjustments vary systematically across vocal frequency bands.

We then wondered whether there is a systematic tendency of hue as a function of frequency. First, we observed that the average hue adjustments followed the same very clear trend across frequencies for the first two stimulus sets of pure tones and the set with the vocal frequency bands: Low frequencies were associated with bluish hues, and with increasing frequencies hues changed continuously through the hue spectrum until they reached yellow hues for high frequencies (see Fig. 3a). To capture this relationship, we calculated Pearson correlations between hue and frequency level. All three datasets produced negative correlation coefficients that were so high that they reached significance [$r(4) = -0.93$, $p = 0.008$; $r(6) = -0.89$, $p = 0.003$; $r(3) = -0.91$, $p = 0.03$] despite the low number of cases ($n = 6$, 8, and 5). In Fig. 3a we pooled the data of the stimulus set by z-scoring the frequency levels to account for range effects. The relationship between the z-scored frequency levels and hue adjustments explained 80% of the variance in the data and was highly significant [$r(17) = -0.89$, $p < 0.001$]. Note that this correlation for the combined datasets is significant even if we apply a Bonferroni correction for testing all five single datasets before calculating the correlation for the combined datasets ($\alpha = 0.05/6$). Hence, these results suggest a clear relationship between levels of frequencies and average hue adjustments.

Similar to our analysis of chroma, we calculated correlation coefficients for each individual observer and tested their difference from zero based on Fisher's *z*-transforms. However, the clear negative relationship between hue and frequency in the average data, is only faint when tested across individuals: the average negative correlation was significant across individuals for the vocal frequency bands [Mr = -0.34, $t(43) = -3.0$, $p = 0.005$, $d = 0.45$], but not for the two sets of pure tones [Mr = -0.17, $t(43) = -1.4$, $p = 0.17$, $d = 0.21$; Mr = -0.10, $t(43) = -1.2$, $p = 0.24$, $d = 0.18$]. When pooling the three datasets (pure sounds set 1, set 2, and vocal frequency bands) the negative correlation just missed significance [Mr = -0.12, $t(43) = -1.9$, $p = 0.058$, $d = 0.29$]. This was the case despite the high number of participants (44), and hence level of statistical power (ranging from 0.9 to 1). These observations suggest that the relationship between hue and frequencies is an overall trend of the aggregated data, but not a common feature of each individual observer's adjustment.

To clarify the role of individual differences, we inspected correlations between frequency levels and hue adjustments for each individual observer. We applied a Bonferonni correction for the 44 individual tests ($\alpha = 0.05/44$). Despite this correction, there were significant correlations for three observers (see Fig. S3 for details). Two of them yielded significant negative correlations as it was the case for the average data (both $p < 0.001$), but one of them yielded a significant positive correlation that contradicts the pattern of the average data [$r(17) = 0.84$, $p < 0.001$]. These observations exemplify systematic individual differences and reinforce the idea that the negative relationship between frequency levels and hue is a feature of the aggregated data rather than a tendency in each individual set of adjustments.

In addition to the relationship between frequencies and hues, we also found some evidence for a relation between hue adjustments and the four levels of loudness of the pure tones in the third stimulus set (cf. Fig. 3b): Increasing the loudness of pure tones made hue change from hues that correspond to short-wavelength spectral lights (blue) towards hues that correspond to spectral lights with longer wavelengths (yellow). The average hue adjustments in Fig. 3b followed a near perfect linear trend across the four loudness levels [$r(2) = -0.996$, $p = 0.004$]. The *t*-test across individual observers reproduced a negative trend that was close to significance [Mr = -0.23, t(43) = -1.8, $p = 0.08$]. Unfortunately, the four stimuli in this set are too few to further investigate correlations for individual observers when also considering Bonferroni corrections across the 44 observers, and they are also too few to draw firm conclusions based on the average data. For this reason, we consider the results about hue-loudness associations to be interesting for further investigation, but yet not conclusive.

However, In contrast to the above results complex sine wave and vocal timbre sounds that varied in their centre of gravity did not yield this pattern at all. Figure 3c shows the average hue adjustments for these two stimulus sets as a function of z-scored frequency levels. Average hue adjustments are concentrated in the yellowish range of hues for all variations in their 'centre of gravity' across both datasets. These two relationships have a very similar profile orientated around yellow hues. This is confirmed by a highly significant positive correlation across the 72 hue steps [$r(70) = 0.55$, $p < 0.001$]. The key difference that might explain this tendency is that both sets of stimuli always had a wide bandwidth (containing frequencies spanning 100–3200 Hz) and always contained high frequencies. While for the other stimuli, the highest frequencies yielded yellow hues, but here, these high frequencies were always present, and hence, always producing yellow hues.

Finally, we also compared median hue adjustments across the six sets of sounds. However, differences just missed significance in the circular median test (P-statistic = 8.9, $p = 0.06$). This result does not allow for a firm conclusion and requires further investigation.

4. Discussion

Through controlling the influence of lightness and defining ways in which timbral sounds can systematically change, evidence is presented for a variety of new correspondences between hearing and vision that have been previously obscured until now. First a linear relationship between the average frequency in a sound and chroma was established across a wide variety of stimuli and contexts. We also observed a positive linear relationship between loudness and chroma (Giannakis, 2001). There were also relationships with hue: for sounds that spanned different frequency ranges (pure tones, vocal frequency bands), low frequencies produced blue hues and frequencies above 800 Hz trended towards yellow hues. Interestingly, the sounds that had a wide bandwidth of frequencies, always produced yellow hues, which may be due to the constant presence of frequencies above 800 Hz. For the first time, we show that this association between frequency and yellow hues operates independently of lightness (Simpson *et al*., 1956; Spence, 2011). We also report a new loudness–hue relationship with quieter sounds yielding bluer hues and louder ones yielding yellower hues. Overall, our investigations reveal specific associations between sound and colour that operate independently of lightness.

The chroma of colours chosen by participants appeared to increase in response to increasing a variety of auditory attributes, namely pitch, loudness, frequency-range and 'centre of gravity.' One important question to ask is whether these correspondences are independent or different expressions of the same fundamental process. While previously these are typically described as pitch–chroma or loudness–chroma correspondences, it could be that both are a result of matching two set of stimuli (e.g., auditory and visual) based on the most obvious perceptual changes (e.g., pitch and chroma), ranked from low to high. This might also explain why some appear to flexibly re-apply themselves in a variety of contexts. These could manifest in lower-level intensity matching, or even higher-level evaluations. Furthermore, this might be a different type of correspondence from those where the presence of an auditory attribute is matched to the presence of a visual attribute, such as with frequencies over 800 Hz and yellow hues.

The current classification of correspondences has focused on their potential aetiology, as well as their effect on low-level perceptual processing and higher-level decision making (Spence, 2011). Other aspects have been under explored, such as to what extent they apply themselves across a variety of contexts and how multiple correspondences interact. For example, eliminating variations in lightness reveal for the first time that the relationship between pitch and chroma is linear, unlike prior studies which reported a quadratic relationship with highest chroma nearest 262 Hz (Ward *et al*., 2006). One explanation for this discrepancy is that stronger pitch–lightness correspondences override pitch–chroma correspondences. If the most perceptually dominant attribute is chosen for matching, then lightness may be primary characteristic chosen for matching, with secondary consideration for other qualities of colour.

4.1. Potential Methodological Problems

Several of our results on sound–hue correspondences were significant for average data, but did not reach significance when tested across individual observers, even despite the higher statistical power due to a higher number of observer data points. Additional analyses suggested that this was due to strong and systematic individual differences in sound–hue associations. However, if there is a central tendency in the association between sound and hue, it should, ultimately, also appear in test across individual differences if statistical power is high enough. To clarify this issue, future investigation could examine whether the sound–hue associations can be replicated in tests across individuals when involving a larger sample of participants to increase statistical power further.

The size of the participant sample becomes particularly important when considering the investigation of gender differences. Our sample of participants included more female (75%) than male observers. In order to make sure that our results were not exclusive to one or the other gender we redid all analyses for women and men separately (see Figs S5–S10 in the online Supplementary Material). Results revealed the same patterns as for the complete sample of participants. However, in a few cases the patterns were not always significant for the group of male observers. This may be explained by the lower number of observers and the resulting lower statistical power. The similarity of the patterns suggests that differences between women and men are small, if they exist at all. Future studies would require larger and more equally distributed samples to tease apart any potential fine-grained differences beyond the strong general tendencies of sound–colour associations observed in our study.

When increasing the number of male observers it also becomes important to screen for colour vision deficiencies (Sharpe *et al*., 1999). Because the most prevalent colour vision deficiencies are X-chromosome-linked they mainly occur in male observers (8–10%), and only very rarely in women (< 1%). Comparing cross-modal correspondences in normal trichromats and dichromats (colour blind) observers in future studies could help to clarify the link between cross-modal correspondences, learned associations, and sensory mechanisms of colour perception (for discussion see Spence, 2011). For example, such an approach has been successful in studying the origin of colour preferences (Álvaro *et al*., 2015).

Another factor that might be interesting to further investigate is the precise measure of lightness. We controlled lightness by holding L* in CIELUV colour space constant, which is equivalent to constant luminance. However, this approach controls a sensory estimate of lightness, but does not completely account for perceived lightness. In particular, perceived lightness might still slightly vary during adjustments due to the Helmholtz–Kohlrausch effect (Nayatani, 1998). This effect indicates that subjective lightness can vary with intense chroma, and this effect is variable across both hue and across individuals (Ayama and Ikeda, 1998). This effect can be minimised by limiting adjustments to a maximum chroma that is equal across hues as in our method (also see Witzel and Franklin, 2014). However this could be double-checked in the future by controlling brightness-luminance ratios through flicker fusion tasks, Ware–Cowan equations or psychophysical data (Fairchild, 1998, *p*. 142; Pridmore, 2007).

4.2. Relative and Absolute Correspondences?

We found that correlations between pitch and chroma flexibly re-apply themselves across a variety of contexts (e.g., 100–3200 Hz or 440–880 Hz). However, saturation towards yellow hues was more rigid, primarily occurring for frequencies above 800 Hz. The different characteristics of flexible and rigid correspondences may indicate different underlying mechanisms. For flexible correspondences, discrete values for the auditory stimulus might be abstracted into simpler representations of magnitude or polarity, where a stimulus is rated as relatively low to high based on where the upper and lower bounds of stimulation are in a given context. One such example of this are the pitch–chroma mappings found to re-apply themselves to different frequency ranges. Walsh (2003) proposes such a mechanism to abstract magnitudes between seemingly independent qualities (time, space and quantity) in the parietal cortex. Of interest to the present research is that disruption of the intraparietal cortex can eliminate cross-modal integration (Bien *et al*., 2012), if flexible correspondences are based on magnitude/polarity matching in the parietal lobe, then it would be predicted that these correspondences would be reduced through parietal disruption similar to disruptions seen to developmental synaesthesia (Esterman *et al*., 2006; Muggleton *et al*., 2007). The fact that pitch-–chroma correspondences appear to be dominated by seemingly stronger pitch–lightness correspondences could suggest a few possibilities. Either lightness is a favoured visual dimension for magnitude/polarity matching, or pitch-–lightness is a rigid correspondence less affected by context (Thornley Head, 2006; Ward *et al*., 2006) and so takes precedence over correspondences that need to be abstracted into magnitudes first.

Rigid audio-visual correspondences appear to be less influenced by their context, with specific auditory characteristics linked to a single visual dimension. For a variety of stimuli, we repeatedly noticed that sounds which had frequencies over 800 Hz were being paired with yellow hues, this was true for pure tones, complex sine wave stimuli and vocal timbre sounds. The mechanisms behind this are unclear at present, they may be based on structural similarities in cortical representation, result from learned correlations with the environment (Spence and Deroy, 2012), or even linked through matching emotional valence with higher frequencies and yellow hues both associated with positive emotions (Palmer *et al*., 2013; although see Schloss *et al*., 2016).

4.3. Comparisons with Synaesthesia

Criteria for the definition of synaesthesia diverge in the literature and are debated (Cohen Kadosh and Terhune, 2012; Eagleman, 2012; Hupé and Dojat, 2015; Simner, 2012ab; Ward, 2013). As a consequence, the precise link between synaesthesia and cross-modal correspondences remains unclear. In their approach, Deroy and Spence (2013) defined criteria to distinguish between cross-modal correspondences and synaesthesia. One of their criteria is that synaesthesia consists of absolute associations between inducers and concurrents, while cross-modal correspondences are relative to the stimulus set. Our observed frequency–chroma correspondences appear to follow this rule, in that they flexibly re-apply themselves to different auditory contexts, as well as displaying a linear relationship in equiluminant contexts, but quadratic in non-equiluminant contexts (Ward *et al*., 2006). While frequency–chroma mappings may be flexible for correspondences, in sound-colour synaesthesia this relationship may be more absolute (Thornley Head, 2006), and hence we would expect sound–colour synaesthetes to maintain a quadratic relationship even if presented with equiluminant colours to best approximate their synaesthetic photisms. Furthermore, this relative/absolute distinction appears to not be true for all audiovisual correspondences. In particular, the selection of yellow hues for sounds that featured frequencies above 800 Hz appeared to be a more absolute association, present in across a variety of contexts. As such, seemingly clear-cut distinctions on first glance may not hold true for all exemplars, and leaves open the question of how different cross-modal correspondences may vary in their similarity to synaesthesia.

The development of synaesthetic links are influenced by associations present during infancy, both in the external environment (Witthoft and Winawer, 2006;

Witthoft *et al*., 2015), and internally, from cross-modal correspondences (Walker *et al*., 2010). Variations in correspondences, in terms of their underlying mechanisms, influence on perceptual processing, or susceptibility to changes in context, may have different levels of influence on the development of synaesthetic links. Sound–colour synaesthesia in particular appears to have more in common with rigid correspondences as they are less affected by changes in context, such as incorrect note naming (Thornley Head, 2006). As such, rigid correspondences (e.g., frequencies above 800 Hz and yellow hues) may have increased influence on synaesthesia; indeed, there are multiple reports showing that synaesthetic photisms trend towards yellow hues when higher frequencies are present in pure tones, complex sine waves, and vocal sounds (Moos *et al*., 2014; Orlandatou, 2012). As such, these more absolute mappings may be a stronger influence on synaesthesia than flexible correspondences (e.g., frequency–chroma) that appear to transition from a linear to a quadratic relationship when variations of luminance are introduced. It appears that any linear frequency–chroma correspondence is dominated by stronger pitch-luminance associations, in both synaesthetes and controls (Ward *et al*., 2006). The mechanisms behind this hierarchy of correspondences and its implication for synaesthesia are a currently evolving area (Jonas *et al*., in press).

4.4. Correspondences and Sensory Substitution

Mapping out the varieties of cross-modal correspondences that exist have important implications for designing intuitive multisensory technology (Hamilton-Fletcher and Ward, 2013). For example, if colour information were to be substituted by audition, pitch–lightness mappings are strongest (Ward *et al*., 2006), implying that the highest and lowest frequencies should be reserved for light and dark colours. This leaves options for chroma to either increase with loudness or richer timbres (Ward *et al*., 2006). Specific hues have been associated with specific timbres, with green-red hues for vowels (Moos *et al*., 2014) and blue-yellow hues for timbral sounds spanning low and high frequency ranges respectively. The use of these intuitive cross-sensory mappings has already been shown to improve performance with colour-to-sound sensory substitution devices (Hamilton-Fletcher *et al*., 2016a). The use of audiovisual correspondences can also enhance the aesthetic and emotive appeal of such technology in both the sighted and the blind (Hamilton-Fletcher *et al*., 2016b, c; Ward *et al*., 2008). Through benefits to function as well as aesthetics and emotive appeal, this can promote a longer-term use of technology. For the blind, long-term adoption of audiovisual sensory substitution technology has led to auditory stimulation producing visual experiences in the blind (Ward and Meijer, 2010), in essence creating a practical form of acquired synaesthesia (Ward and Wright, 2014). While the present findings illustrate some of the ways visual and auditory dimensions are associated in sighted individuals, understanding the role that visual experience plays in correspondences has large implications for how specific correspondences operate and are altered by visual deprivation (Deroy *et al*., 2016; Fryer *et al*., 2014). These findings in turn also have important implications for creating intuitive visual-assistive devices across different user groups, such as the early-blind, late-blind, partially-sighted and those transitioning into sight loss (Hamilton-Fletcher *et al*., 2016c).

4..5. Conclusion

The relationship between sound and colour has inspired discussion examining their structures in physics (Newton, 1979), their aesthetic appeal in art (Jewanski, 2010) and their processing in psychology (Spence, 2011). By examining our correspondences between hearing and vision, we can improve our understanding of the underlying processes involved in multisensory perception. The use of equiluminant conditions in the present research have illustrated previously hidden correspondences between sound and chroma/hue as well as disentangled the influence of lightness in explanations of these correspondences. Establishing that there is variation in which sound-colour correspondences appear to rigidly follow certain auditory features or flexibly re-apply themselves to new auditory contexts reveals another way in which correspondences differ from one another. This in turn helps further cast light on the underlying mechanisms that give rise to cross-modal correspondences.

Acknowledgements

The research was supported by a studentship from the ESRC, the University of Sussex and the RM Phillips Foundation. CW was supported by a German Academic Exchange Service (DAAD) postdoctoral fellowship and by the grant ‘Cardinal Mechanisms of Perception’ No SFB TRR 135 from the Deutsche Forschungsgemeinschaft.

References

Afra, P., Funke, M. and Matsuo, F. (2009). Acquired auditory-visual synesthesia: A window to early cross-modal sensory interactions*, Psychol. Res. Behav. Manag.* **2**, 31–37.

Álvaro, L., Moreira, H., Lillo, J. and Franklin, A. (2015). Color preference in red–green dichromats*, Proc. Natl Acad Sci.* **112**, 9316–9321.

Ayama, M. and Ikeda, M. (1998). Brightness-to-luminance ratio of colored light in the entire chromaticity diagram*, Color Res. Appl.* **23**, 274–287.

Batschelet, E. (1981). *Circular Statistics in Biology,* Academic Press, London, UK.

Berens, P. (2009). CircStat: A MATLAB toolbox for circular statistics*. J. Stat. Softw.* **31**, 1-21.

Bien, N., Ten Oever, S., Goebel, R. and Sack, A. T. (2012). The sound of size: Crossmodal binding in pitch-size synesthesia: A combined TMS, EEG and psychophysics study*, Neuroimage* **59**, 663–672.

Boersma, P. and Weenink, D. (2012). *Praat: Doing Phonetics by Computer*. Retrieved from http://www.praat.org/.

Bond, B. and Stevens, S. (1969). Cross-modality matching of brightness to loudness by 5-year-olds, *Percept. Psychophys*. **6**, 337–339.

Braaten, R. (1993). *Synesthetic Correspondence between Visual Location and Auditory Pitch in Infants*. Paper presented at the 34th Annual Meeting of the Psychonomic Society, Washington, DC, USA.

Brainard, D. H. (1997). The Psychophysics Toolbox, *Spat. Vis.* **10**, 433–436.

Cohen Kadosh, R. and Terhune, D. B. (2012). Redefining synaesthesia? *Br. J. Psychol*. **103**, 20–23.

Deroy, O., & Spence, C. (2013). Why we are not all synesthetes (not even weakly so). Psychonomic Bulletin & Review, 20(4), 643-664.

Deroy, O., Fasiello, I., Hayward, *V*. and Auvray, M. 2016. Differentiated audio-tactile correspondences in sighted and blind individuals*, J. Exp. Psychol. Hum. Percept. Perform.* **42**, 1204–1214.

Dolscheid, S., Shayan, S., Majid, A. and Casasanto, D. (2013). The thickness of musical pitch psychophysical evidence for linguistic relativity*, Psychol. Sci.* **24**, 613–621.

Eagleman, D. M. (2012). Synaesthesia in its protean guises, *Br. J. Psychol*. **103**, 16–19.

Esterman, M., Verstynen, T., Ivry, R. B. and Robertson, L. C. (2006). Coming unbound: Disrupting automatic integration of synesthetic color and graphemes by transcranial magnetic stimulation of the right parietal lobe*, J. Cogn. Neurosci.* **18**, 1570–1576.

Evans, K. K. and Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features*, J. Vis.* **10**, 6. doi: 10.1167/10.1.6

Fairchild, M. D. (1998). *Color Appearance Models*. Addison Wesley Longman, Reading, MA, USA.

Fisher, R. A. (1921). On the 'probable error' of a coefficient of correlation deduced from a small sample, *Metron* **1**, 3–32.

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques*, J. Acoust. Soc. Am.* **102**, 1213–1222.

Fletcher, H. and Munson, W. A. (1933). Loudness, its definition, measurement and calculation*, Bell Syst. Techn. J.* **12**, 377–430.

Forder, L., He, X., Witzel, C. and Franklin, A. (2014). Speakers of different colour lexicons differ only in post-perceptual processing of colour, *Perception* (ECVP Abstract Suppl.) **43**, 33.

Fryer, L., Freeman, J. and Pring, L. (2014). Touching words is not enough: How visual experience influences haptic–auditory associations in the 'Bouba–Kiki' effect*, Cognition* **132**, 164–173.

Giannakis, K. (2001). Sound mosaics: A graphical user interface for sound synthesis based on audio-visual associations. Doctoral dissertation, Middlesex University, UK. Retrieved from http://eprints.mdx.ac.uk/6634/1/Konstantinos-sound_mosaics.phd.pdf

Goodyear, B. G. and Menon, R. S. (1998). Effect of luminance contrast on BOLD fMRI response in human primary visual areas*, J. Neurophysiol.* **79**, 2204–2207.

Hamilton-Fletcher, G. and Ward, J. (2013). Representing colour through hearing and touch in sensory substitution devices*, Multisens. Res.* **26**, 503–532.

Hamilton-Fletcher, G., Wright, T. D. and Ward, J. (2016a). Cross-modal correspondences enhance performance on a colour-to-sound sensory substitution device*, Multisens. Res.* **29**, 337–363.

Hamilton-Fletcher, G., Mengucci, M. and Medeiros, F. (2016b). Synaestheatre: Sonification of coloured objects in space, in: *Proceedings of the International Conference on Live Interfaces 2016,* Sussex, Brighton, UK, pp 252–256*.*

Hamilton-Fletcher, G., Obrist, M., Watten, P., Mengucci, M. and Ward, J. (2016c). 'I always wanted to see the night sky': Blind user preferences for Sensory Substitution Devices, in: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose, CA, USA, pp. 2162–2174.

Hupé, J. M. and Dojat, M. (2015). A critical review of the neuroimaging literature on synesthesia*, Front. Hum. Neurosci.* **9**, 103. doi: 10.3389/fnhum.2015.00103

ISO (2003). *Acoustics – Normal Equal-Loudness Level Contours.* International Organisation for Standardisation, Geneva, Switzerland.

Jewanski, J. (2010). Color-tone analogies: A systematic presentation of the principles of correspondence, in: *Audiovisuology: Compendium. A Multidisciplinary Survey of Audiovisual Culture,* D. Daniels, S. Naumann and J. Thoben (Eds), pp. 338–347, König, Köln, Germany.

Jonas, C., Spiller, M. J. and Hibbard, P. (in press). Summation of visual attributes in auditory-visual crossmodal correspondences*, Psychonom. Bull. Rev*.

Ludwig, *V*. U. and Simner, J. (2013). What colour does that *feel*? Tactile-visual mapping and the development of cross-modality, *Cortex* **49**, 1089–1099.

Ludwig, *V*. U., Adachi, I. and Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (*Pan troglodytes*) and humans*, Proc. Natl Acad Sci.***108**, 20661–20665.

Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness*, Am. J. Psychol.* **87**, 173–188.

Marks, L. E. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination*, J. Exp. Psychol. Hum. Percept. Perform.* **13**, 384–394.

Martino, G. and Marks, L. E. (1999). Perceptual and linguistic interactions in speeded classification: Tests of the semantic coding hypothesis, *Perception* ***28***, 903–923.

Melara, R. D. (1989). Dimensional interaction between color and pitch, *J. Exp. Psychol. Hum. Percept. Perform.* **15**, 69–79.

Mondloch, C. J. and Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children, *Cogn. Affect. Behav. Neurosci.* **4**, 133–136.

Moos, A., Smith, R., Miller, S. R. and Simmons, D. R. (2014). Cross-modal associations in synaesthesia: Vowel colours in the ear of the beholder, *Iperception* **5**, 132–142.

Muggleton, N., Tsakanikos, E., Walsh, *V*. and Ward, J. (2007). Disruption of synaesthesia following TMS of the right posterior parietal cortex, *Neuropsychologia* **45**, 1582–1585.

Mulert, C., Jäger, L., Propp, S., Karch, S., Störmann, S., Pogarell, O., Möller, H-J. and Hegerl, U. (2005). Sound level dependence of the primary auditory cortex: Simultaneous measurement with 61-channel EEG and fMRI, *Neuroimage* **28**, 49–58.

Nayatani, Y. (1998). A colorimetric explanation of the Helmholtz–Kohlrausch effect, *Color Res. Appl.* **23**, 374–378.

Newton, I. (1979). *Opticks: or a Treatise of the Reflections, Refractions, Inflections and Colours of Light*. Courier Dover Publications, New York, NY, USA.

Novich, S., Cheng, S. and Eagleman, D. M. (2011). Is synaesthesia one condition or many? A large-scale analysis reveals subgroups, *J. Neuropsychol.* **5**, 353–371.

Orlandatou, K. (2012). The role of pitch and timbre in the synaesthetic experience, in: *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music,* Thessaloniki, Greece, pp. 751–758.

Palmer, S. E., Schloss, K. B., Xu, Z. and Prado-León, L. R. (2013). Music–color associations are mediated by emotion, *Proc. Natl Acad. Sci. USA* **110**, 8836–8841

Parise, C. *V*., Knorre, K. and Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing, *Proc. Natl Acad Sci. USA* **111**, 6104–6108.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies, *Spat. Vis*. **10**, 437–442.

Pridmore, R. W. (2007). Effects of luminance, wavelength and purity on the color attributes: Brief review with new data and perspectives, *Color Res. Appl.* **32**, 208–222.

Root, R. T. and Ross, S. (1965). Further validation of subjective scales for loudness and brightness by means of cross-modality matching, *Am. J. Psychol.* **78**, 285-289.

Schloss, K., Lai, Y.-H. and Witzel, C. (2016). Yellow is no happier than blue when lightness and chroma are controlled, *J. Vis.* **16**, 624. doi: 10.1167/16.12.624

Sebba, R. (1991). Structural correspondence between music and color, *Color Res. Appl.* **16**, 81–88.

Sharpe, L. T., Stockman, A., Jägle, H. and Nathans, J. (1999). Opsin genes, cone photopigments, color vision, and color blindness. In *Color Vision: From Genes to Perception*, K. R. Gegenfurtner and L. T. Sharpe (Eds), pp. 3–51, Cambridge University Press, Cambridge, UK.

Simner, J. (2012a). Defining synaesthesia, *Br. J. Psychol*. **103**, 1–15.

Simner, J. (2012b). Defining synaesthesia: A response to two excellent commentaries, *Br. J. Psychol*. **103**, 24–27.

Simner, J. (2013). The ‘rules’ of synesthesia, in: *Oxford Handbook of Synesthesia*, J. Simner and E. Hubbard (Eds), pp. 149–164. Oxford University Press, Oxford, UK.

Simner, J. and Ludwig, *V*. U. (2012). The color of touch: A case of tactile-visual synaesthesia, *Neurocase* **18**, 167–180.

Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S. A., Fraser, C., Scott, K. and Ward, J. (2006). Synaesthesia: The prevalence of atypical cross-modal experiences*, Perception* **35**, 1024–1033.

Simpson, R. H., Quinn, M. and Ausubel, D. P. (1956). Synesthesia in children: Association of colors with pure tone frequencies*, J. Genet. Psychol.* **89**, 95–103.

Spence, C. (2011). Crossmodal correspondences: a tutorial review*, Attent. Percept. Psychophys.* **73**, 971–995.

Spence, C. and Deroy, O. (2012). Crossmodal correspondences: Innate or learned? *Iperception* **3**, 316. doi: 10.1068/i0526ic

Stevens, J. C. and Marks, L. E. (1965). Cross-modality matching of brightness and loudness*, Proc. Natl Acad. Sci. USA* **54**, 407–411.

Thornley Head, P. D. (2006). Synaesthesia: Pitch-colour isomorphism in RGB-space? *Cortex* **42**, 164–174.

Walker, P. (2012). Cross-sensory correspondences and cross talk between dimensions of connotative meaning: Visual angularity is hard, high-pitched, and bright*, Attent. Percept. Psychophys.* **74**, 1792–1809.

Walker, P., Bremner, G., Mason, U., Spring, J., Mattock, K., Slater, A. and Johnson, S. (2010). Preverbal infants’ sensitivity to synaesthetic cross-modality correspondences*, Psychol. Sci.* **21**, 21-25.

Walsh, *V*. (2003). A theory of magnitude: common cortical metrics of time, space and quantity, *Trends Cogn. Sci.* **7**, 483–488.

Ward, J. (2013). Synesthesia, *Annu. Rev. Psychol.* **64**, 49-75.

Ward, J. and Meijer, P. (2010). Visual experiences in the blind induced by an auditory sensory substitution device, *Consc. Cogn.* **19**, 492–500.

Ward, J. and Wright, T. (2014). Sensory substitution as an artificially acquired synaesthesia, *Neurosci. Biobehav. Rev.* **41**, 26–35.

Ward, J., Huckstep, B. and Tsakanikos, E. (2006). Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex* **42**, 264–280.

Ward, J., Moore, S., Thompson-Lake, D., Salih, S. and Beck, B. (2008). The aesthetic appeal of auditory – Visual synaesthetic perceptions in people without synaesthesia*, Perception* **37**, 1285–1296.

Witthoft, N. and Winawer, J. (2006). Synesthetic colors determined by having colored refrigerator magnets in childhood, *Cortex* **42**, 175–183.

Witthoft, N., Winawer, J. and Eagleman, D. M. (2015). Prevalence of learned grapheme-color pairings in a large online sample of synesthetes, *PloS One* **10**, e0118996. doi: 10.1371/journal.pone.0118996

Witzel, C. and Franklin, A. (2014). Do focal colors look particularly ‘colorful’? *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **31**, A365–374.

Supplemental Materials

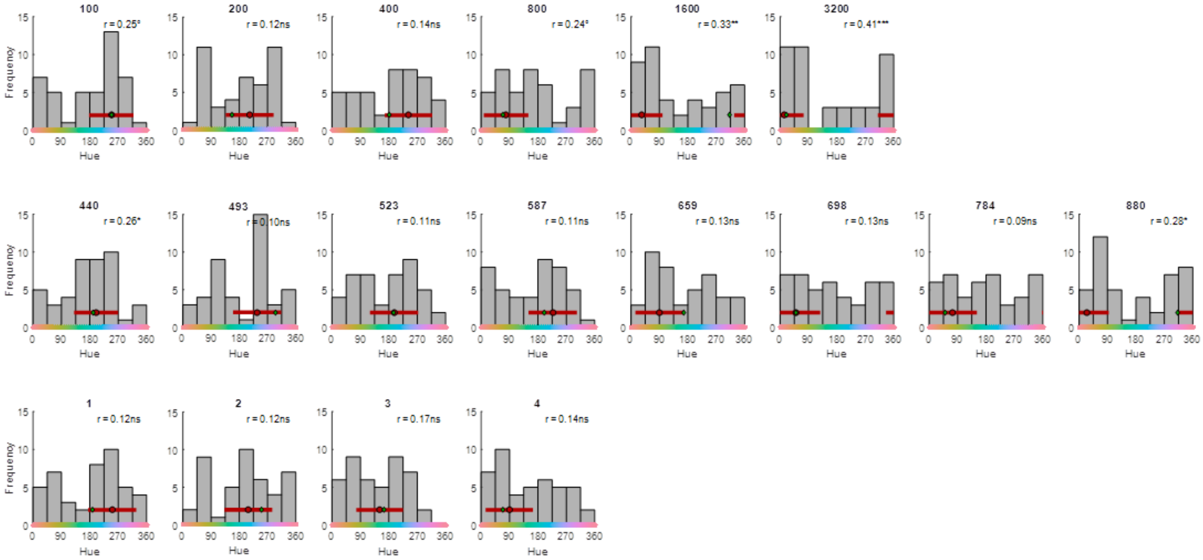

Figure S1. Distribution of hue adjustments for pure tones. The first row (panels a–f) show histograms of hue adjustments for the first set of six pure tones that varied in frequencies; the second row (g–n) displays the histograms for the second set of eight pure tones of varying frequencies; and the last row those for the set that varied in loudness. The bars of the histogram show the frequencies of adjusting a particular hue. For illustration purposes only, they are binned in 45 deg steps (i.e., all statistical analyses are based on the original 5 deg resolution of adjustments). The red disk and bar report the circular mean and standard deviation, the green diamond the circular median of the adjusted hues. Results of Rayleigh's test for uniformity are provided in the upper right corner, where r is the mean vector length and p the probability that the distribution is uniform. ° $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *Note that, according to the Rayleigh tests, the hue adjustments are unimodal for several pure sounds (1600 Hz, 3200 Hz, 440 Hz, 880 Hz) in the first and second set (first and second row).*

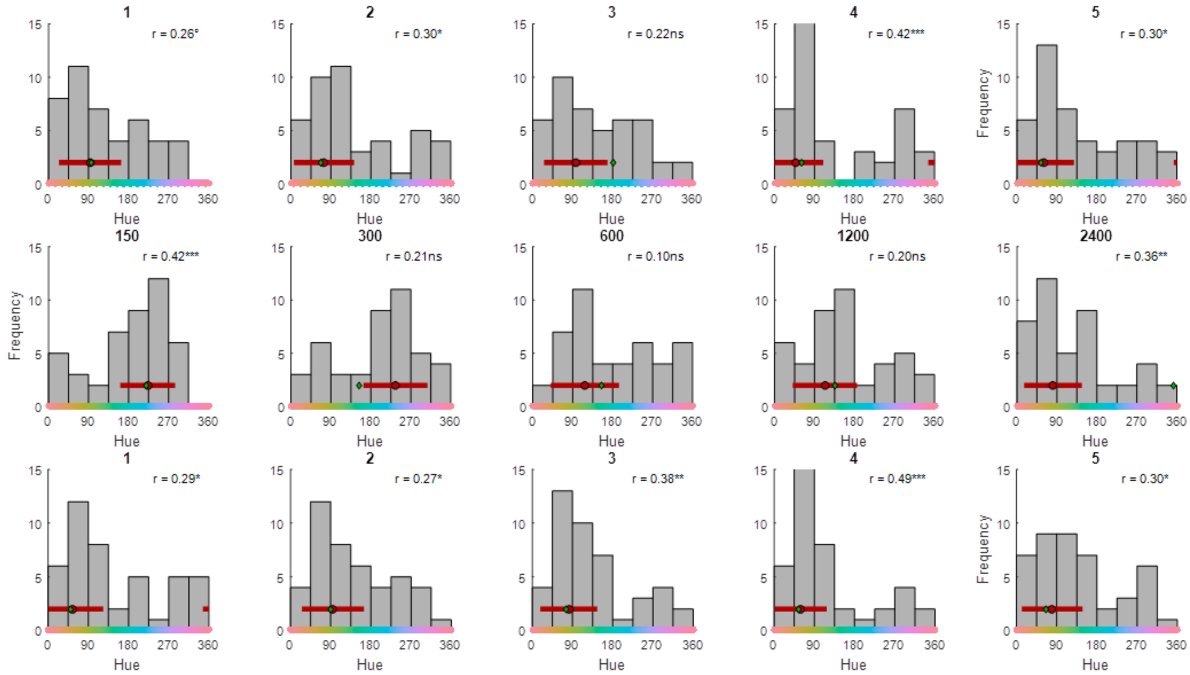

Figure S2. Distribution of hue adjustments for complex sine wave and vocal timbre sounds. Same format as Fig. S1. The first row (panels a–e) illustrates results for complex sine wave sounds varying in their 'centre of gravity' from 1 (lowest) to 5 (highest); the second row (f–j) for vocal frequency bands with their average frequency; and the third row (k–o) those for vocal sounds varying in 'centre of gravity' from 1 (lowest) to 5 (highest). *Note that almost all centre-of-gravity-modulated sounds (first and third row) showed clear unimodal tendencies towards one particular hue, mostly in the yellowish region of the hue spectrum.*

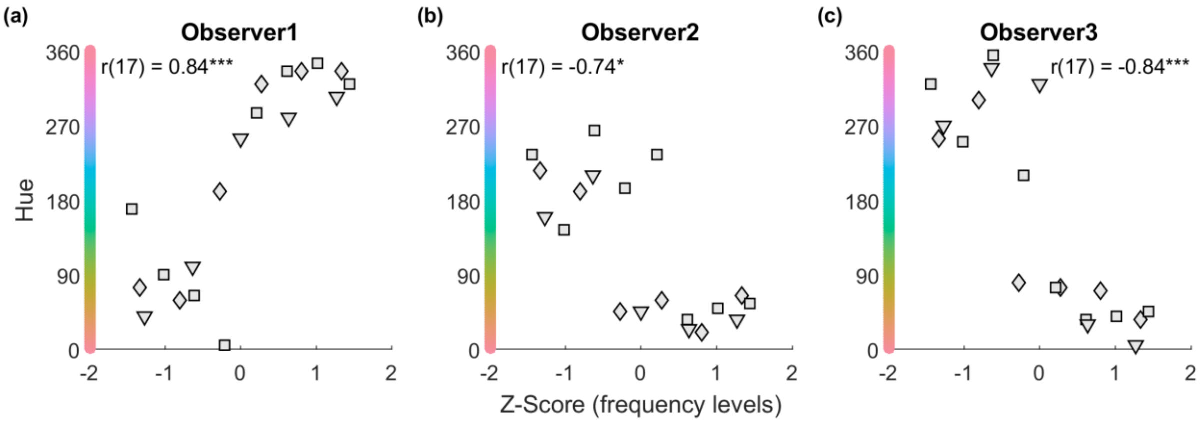

Figure S3. Illustration of individual differences. Format as in Fig. 3a of the main article.

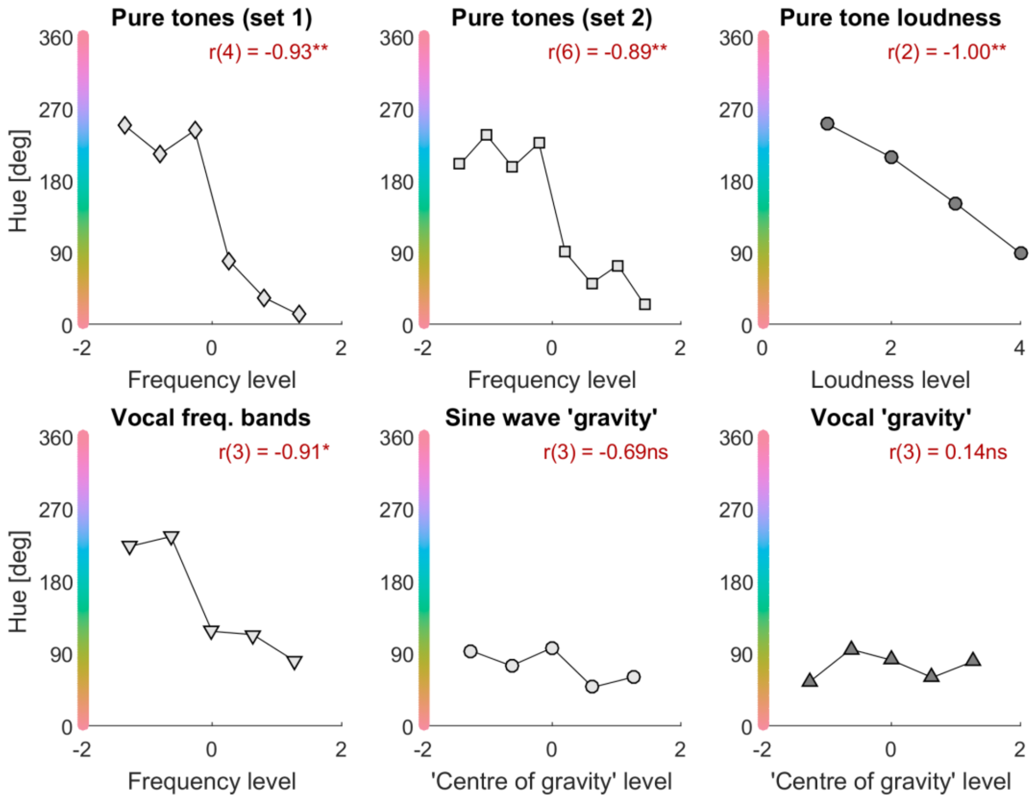

Figure S4. Hue per dataset. Format as in Fig. 3a of the main article.

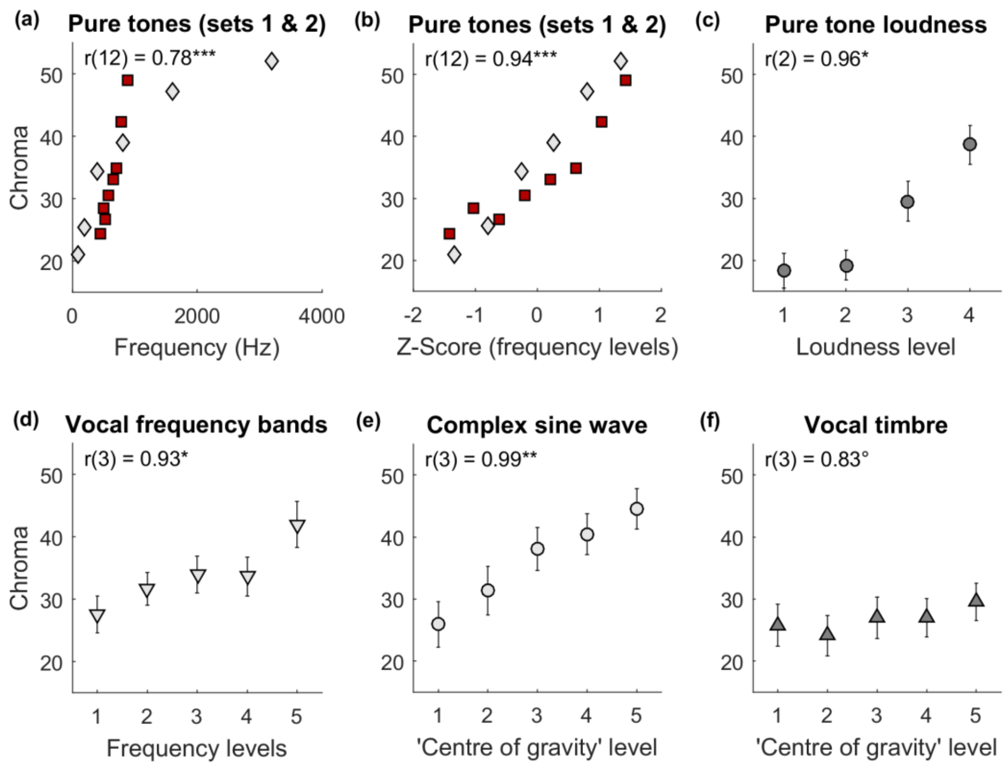

Figure S5. Chroma sound associations per stimulus set in women. Format as in Fig. 1 of the main article.

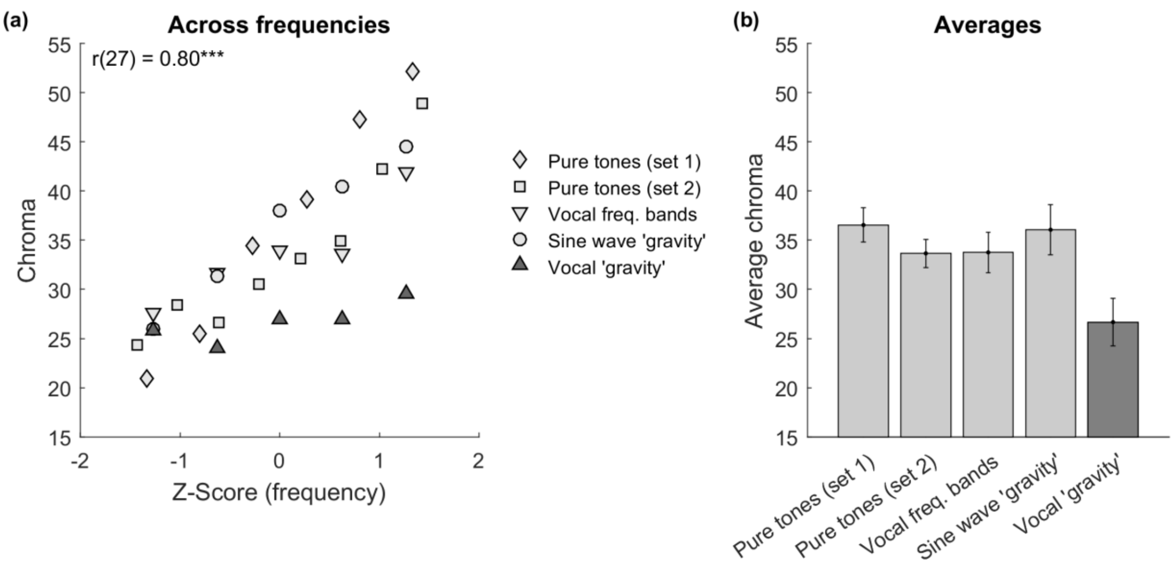

Figure S6. Chroma-sound associations across stimulus sets in women. Format as in Fig. 2 of the main article.

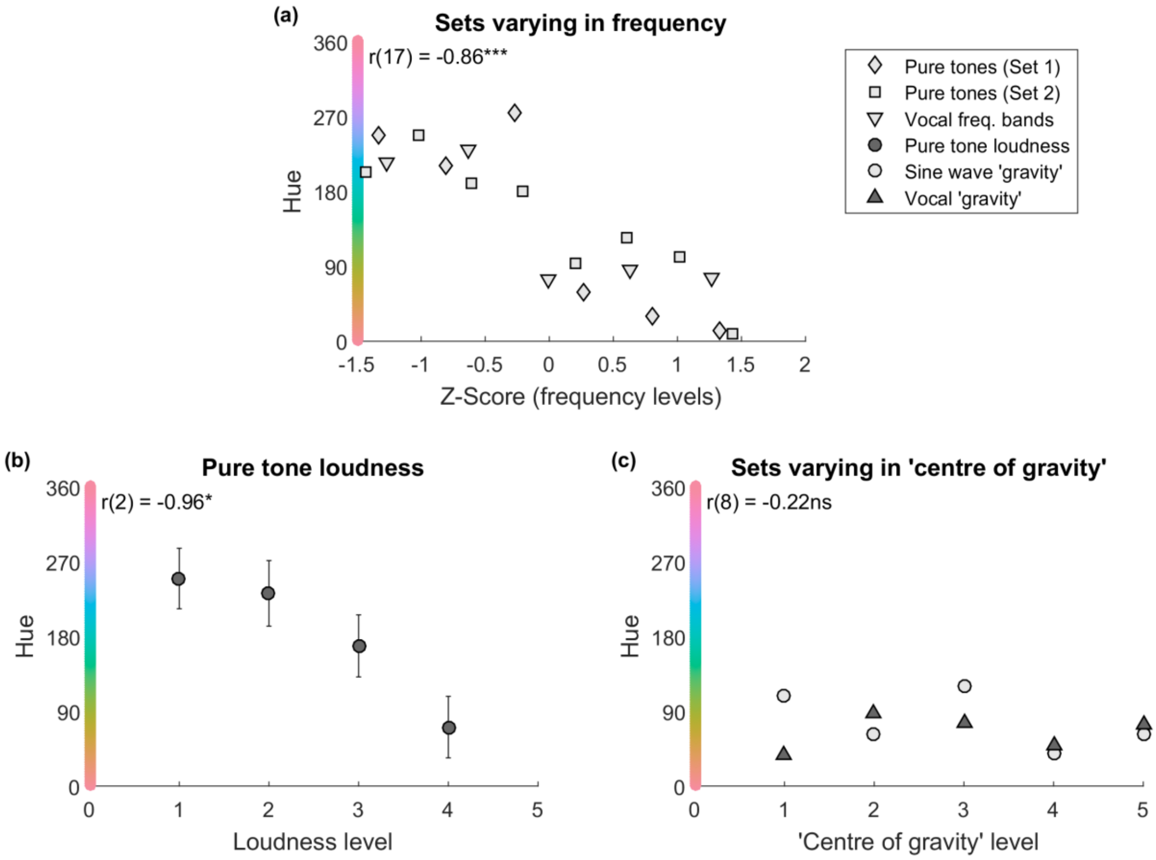

Figure S7. Hue sound associations in women. Format as in Fig. 3 of the main article.

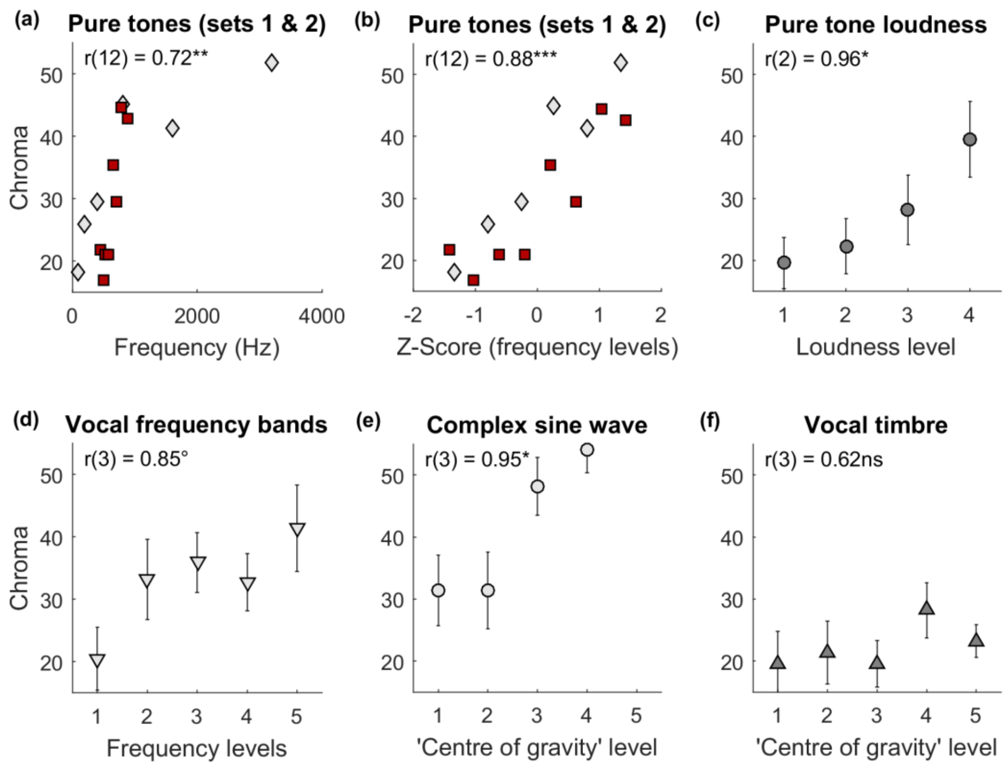

Figure S8. Chroma sound associations per stimulus set in men. Format as in Fig. 1 of the main article.

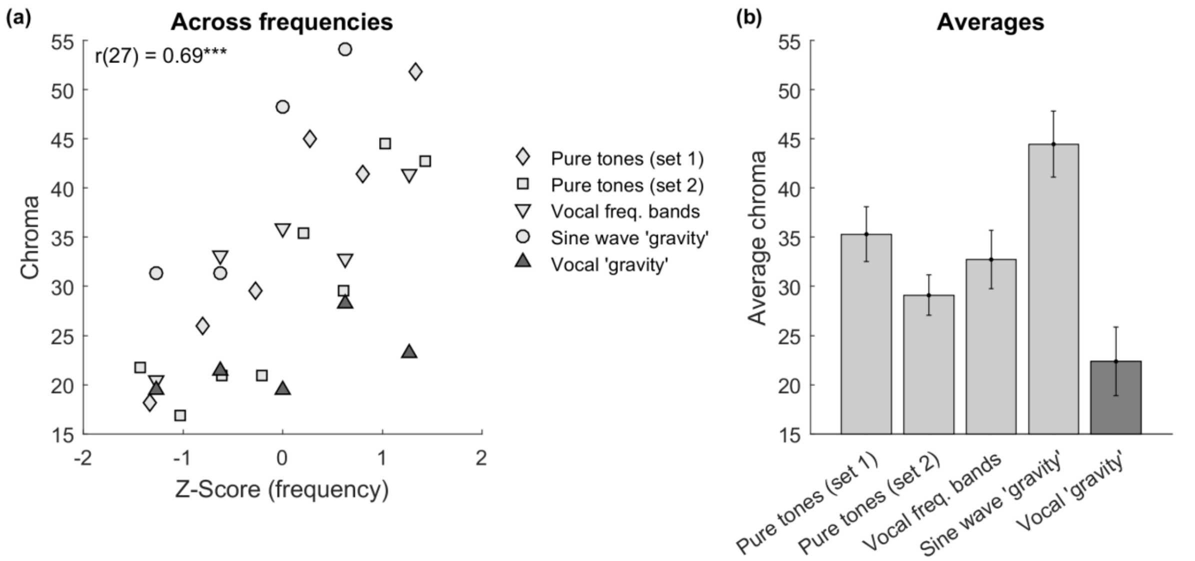

Figure S9. Chroma-sound associations across stimulus sets in men. Format as in Fig. 2 of the main article.

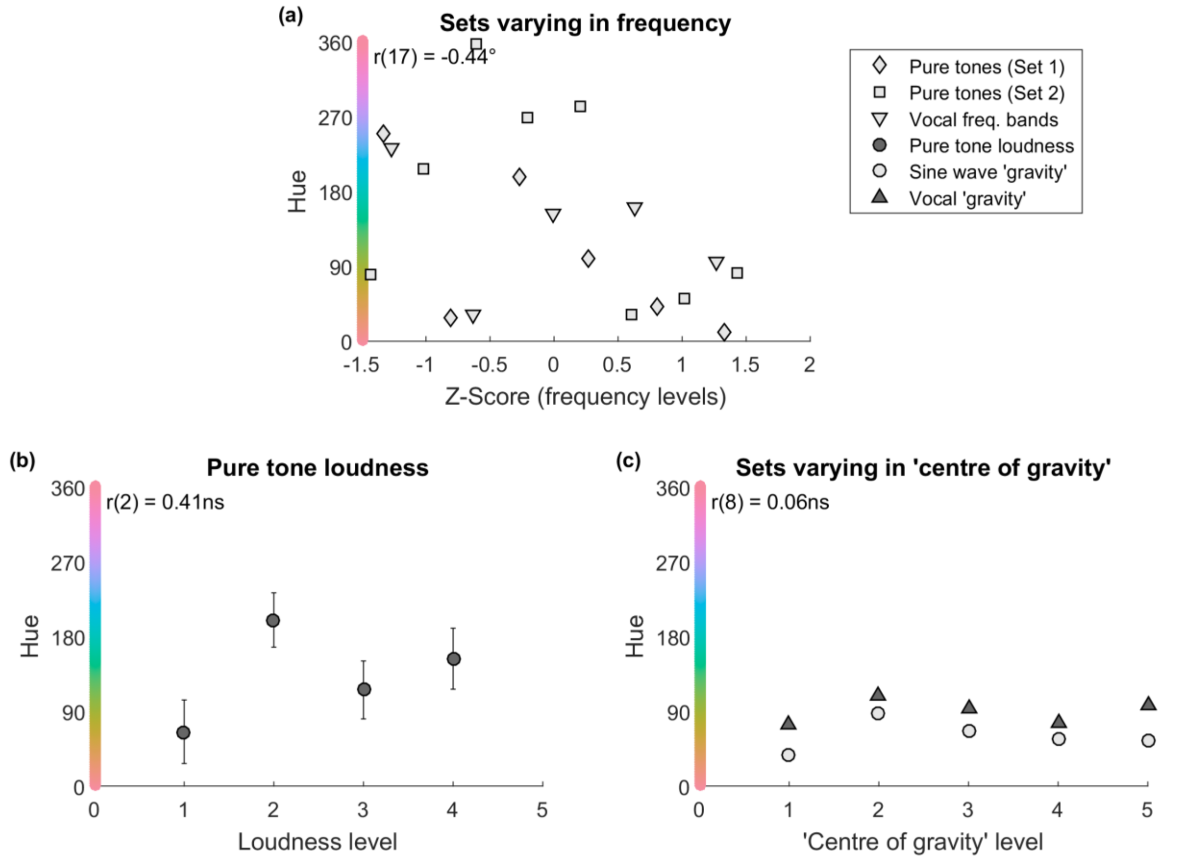

Figure S10. Hue sound associations in men. Format as in Fig. 3 of the main article.